JOURNAL OF COMBINATORIAL THEORY 7, 215–220 (1969)

A Class of Covers for Finite Projective Geometries Which Are Related to the Design of Combinatorial Filing Systems*

GARY G. KOCH

University of North Carolina, Chapel Hill, North Carolina 27514

Communicated by R. C. Bose

Received Jebruary 7, 1968

ABSTRACT

A method is given for selecting a certain subset of m-flats from a finite projective geometry $PG(N, q)$ which cover all $(t-1)$ flats there, where $m \geqslant t-1$. These results have application to the problem of designing efficient information retrieval systems.

Recently there has been some interest in applying certain methods of combinatorial mathematics to the problem of designing filing systems which permit efficient information retrieval [see 1 and 6]. The structure of covers of finite projective geometries becomes of interest here because of the following correspondence.

The basic elements of the file are called *records*. Each record is uniquely identified by an *accession number*. Suppose the data on any record indicates the "presence" or "absence" of each of v attributes. Retrieval will be related to the concept of "presence" and will be expressed in terms of *queries* which specify a subset of attributes whose presence is jointly required. For example, if A_i $(i = 1, 2,..., v)$ denotes the i-th attribute, then the query $A = (A_1, A_2)$ seeks all records which contain both A_1 and A_2. The structure of the filing system is based upon defining certain disjoint subsets (called *buckets*) of a computer memory such that, to each query of interest, there corresponds exactly one bucket. However, the same bucket may be relevant to many queries. Finally, the accession number of any given record is stored exactly once at an address in each of the buckets for which it satisfies at least one of the corresponding queries.

* This research was supported by National Science Foundation Grant GP-5790, Air Force Office of Scientific Research Grant AF-AFOSR-68-1415, and National Institutes of Health, National Institute of General Medical Sciences Grant GM-12868.

Let the v attributes be identified in a one-one way with a subset of the points of the finite projective geometry $PG(N, q)$; i.e.,

$$v \leqslant (q^{N+1} - 1)/(q - 1).$$

Suppose that any given query will involve at most t attributes. Hence, any query will determine at most a $(t - 1)$-flat in the geometry. Let a class of m-flats where $m \geqslant (t - 1)$ be identified with the buckets of the filing system. In order for each query to have a corresponding bucket, this class must be such that every $(t - 1)$-flat in $PG(N, q)$ is contained in at least one of the m-flats. If b m-flats are required, then the resulting construction will be called a (b, t, m) cover. In order for the filing system to be efficient, it is desirable that b be as small as possible.

Simple constructions of (b, t, m)-covers with reasonably small b are not readily available except for some special cases, which will be cited later. Here, we will consider an algorithm for generating a system of covers for $PG(N, q)$ from covers available for $PG(N', q)$ where $N' < N$. No claim of optimality is made. However, because of their basic simplicity they are felt to be of some interest.

LEMMA 1. *There exists a $(b, 1, 1)$-cover for the geometry $PG(N, q)$ with $b = \Phi(N - 1, 0, q)$ where q is a prime power, where*

$$\Phi(N, 0, q) = \frac{q^{N+1} - 1}{q - 1}.$$

PROOF: Consider the set of $b = \Phi(N - 1, 0, q)$ lines through some fixed point P_0 in $PG(N, q)$. This represents a $(b, 1, 1)$-cover since each point of $PG(N, q)$ lies on exactly one of the lines through P_0.

COROLLARY. *There exists an optimal $(b, 1, 1)$-cover for $PG(2, q)$ where q is a prime power.*

PROOF: Since each line in $PG(2, q)$ contains $(q + 1)$ points, we have that

$$b \geqslant (q^2 + q + 1)/(q + 1) > q;$$

i.e., $b \geqslant (q + 1)$. But the construction from Lemma 1 has

$$b = \Phi(1, 0, q) = q + 1$$

and hence is optimal.

LEMMA 2. *There exists a $(b, 1, m)$-cover for the geometry $PG(N, q)$ with $b = \Phi(N - m, 0, q)$ where q is a prime power and $N \geqslant m$.*

PROOF: Consider the set of $b = \Phi(N - m, 0, q)$ m-flats through some fixed $(m - 1)$-flat π_0 in $PG(N, q)$. This represents a $(b, 1, m)$-cover since each point of $PG(N, q)$ which is not on π_0 determines a unique m-flat through π_0.

COROLLARY. *There exists an optimal $(b, 1, m)$-cover for $PG(m + 1, q)$ where q is a prime power; i.e., $b = (q + 1)$.*

PROOF: Since each m-flat in $PG(m + 1, q)$ contains $(q^{m+1} - 1)/(q - 1)$ points, we have that

$$b \geqslant (q^{m+2} - 1)/(q^{m+1} - 1) > q;$$

i.e., $b \geqslant q + 1$. But the construction of Lemma 2 has

$$b = \Phi(1, 0, q) = q + 1.$$

LEMMA 3. *There exists a $(b, 2, 2)$-cover for the geometry $PG(3, q)$ with $b = q^2 + q + 1$ where q is a prime power.*

PROOF: Consider all lines in a given plane π_0 through a given point P_0. There are $(q + 1)$ lines in π_0 through P_0. Form the $(b, 2, 2)$-cover by taking all planes through these lines other than π_0 together with π_0. Hence, $b = q(q + 1) + 1 = q^2 + q + 1$. Since any line L, not in π_0, intersects π_0 in a point P_1 and since the line connecting P_0 and P_1 and the line L determine a unique plane, each line is covered.

LEMMA 4. *There exists a $(b, 2, 2)$-cover for the geometry $PG(N, q)$ with*

$$b = \sum_{\alpha=0}^{N-2} q^\alpha \Phi(\alpha, 0, q),$$

where q is a prime power and $N \geqslant 2$.

PROOF: Consider all lines, in a given $(N - 1)$-flat π_0, through a given point P_0. There are $\Phi(N - 2, 0, q)$ such lines. Form the $(b, 2, 2)$-cover of $PG(N, q)$ by taking all planes through these lines other than those lying in π_0 together with a $(b, 2, 2)$-cover of $PG(N - 1, q)$ as represented by π_0. Since any line L, not in π_0, intersects π_0 in a point P_1 and since the line connecting P_0 and P_1 and the line L determine a unique plane, each line not in π_0 is covered. Those lines in π_0 are covered by the $(b, 2, 2)$-cover developed for $PG(N - 1, q)$ which may be assumed to exist by mathematical induction because of the existence of the $(b, 2, 2)$-cover of $PG(3, q)$ given in Lemma 3. The value of b is determined by noting that through each of the $\Phi(N - 2, 0, q)$ lines through P_0 in π_0, there pass $\{\Phi(N-2, 0, q) - \Phi(N-3, 0, q)\}$ planes not lying in π_0. Thus, b is equal to the sum of $q^{N-2}\Phi(N-2, 0, q)$ and the number of planes needed to cover the $(N-1)$-flat π_0. Proceeding backwards in a recursive manner, we have

$$b = \sum_{\alpha=0}^{N-2} q^{\alpha}\Phi(\alpha, 0, q).$$

LEMMA 5. *There exists a* $(b, 2, m)$*-cover for the geometry* $PG(N, q)$ *with*

$$b = \sum_{\alpha=0}^{N-m} q^{\alpha}\Phi(\alpha, 0, q),$$

where q *is a prime power and* $N \geqslant m \geqslant 2$.

PROOF: Consider all $(m-1)$-flats, in a given $(N-1)$-flat π_0, through some fixed $(m-2)$-flat π_1. There are $\Phi(N-m, 0, q)$ such $(m-1)$-flats. Form the $(b, 2, m)$-cover of $PG(N, q)$ by taking all m-flats through these $(m-1)$-flats other than those lying in π_0 together with a $(b, 2, m)$-cover of $PG(N-1, q)$ as represented by π_0. Since any line L, not in π_0, intersects π_0 in a unique point P_1 which lies in an $(m-1)$-flat π such that $\pi_1 \subseteq \pi \subseteq \pi_0$ and, since L and π determine a unique m-flat, each line not in π_0 is covered. Those lines belonging to π_0 are covered by the $(b, 2, m)$-cover similarly developed for $PG(N-1, q)$, the existence of which may be assumed by mathematical induction since a $(b, 2, m)$-cover exists for $PG(m, q)$. In particular, for the case $N = m$, $b = 1$; for the case $N = m + 1$, $b = q(q + 1)$; and in general b is the sum of

$$\Phi(N-m, 0, q)\{\Phi(N-m, 0, q) - \Phi(N-m-1, 0, q)\}$$

and the number of m-flats needed to cover the $(N-1)$-flat π_0. Proceeding recursively, we have

$$b = \sum_{\alpha=0}^{N-m} q^{\alpha}\Phi(\alpha, 0, q).$$

THEOREM. *There exists a* (b, t, m)*-cover for the geometry* $PG(N, q)$ *with*

$$b = \sum_{\alpha_t=0}^{N-m} \sum_{\alpha_{t-1}=0}^{\alpha_t} \cdots \sum_{\alpha_1=0}^{\alpha_2} q^{\alpha_1+\alpha_2+\cdots+\alpha_t}\Phi(\alpha_1, 0, q),$$

where q *is a prime power and* $N \geqslant m \geqslant t$.

PROOF: Consider the $(m-1)$-flats belonging to a $(b, t-1, m-1)$-cover for the geometry $PG(N-1, q)$ as represented by an $(N-1)$-flat π_0.

This may be assumed by induction. Form the (b, t, m)-cover of $PG(N, q)$ by taking all m-flats through these $(m-1)$-flats other than those lying in π_0 together with a (b, t, m)-cover of $PG(N-1, q)$ as represented by π_0. Since any t-flat π, not in π_0, intersects π_0 in a unique $(t-1)$-flat lying in an $(m-1)$-flat π_1 which belongs to the $(b, t-1, m-1)$-cover in π_0 defined above and, since π and π_1 determine a unique m-flat through π_1, each t-flat not in π_0 is covered. Those t-flats belonging to π_0 are covered by the (b, t, m)-cover similarly developed for $PG(N-1, q)$, the existence of which may be assumed by induction because of Lemmas 1–5. In these constructions b is the sum of

$$\left\{\sum_{\alpha_{t-1}=0}^{N-m} \sum_{\alpha_{t-2}=0}^{\alpha_{t-1}} \cdots \sum_{\alpha_1=0}^{\alpha_2} q^{\alpha_1+\alpha_2+\cdots+\alpha_{t-1}} \Phi(\alpha_1, 0, q)\right\}$$
$$\times \{\Phi(N-m, 0, q) - \Phi(N-m-1, 0, q)\}$$

and the number of m-flats needed to cover the $(N-1)$-flat π_0. Proceeding recursively, we have the stated value of b.

As can be seen from the sequence of results, the induction is first with respect to N, then with respect to m, and finally with respect to t. Finally, it should be mentioned that there do exist covers which are much better than what has been outlined here. These are based upon the concept of spread as discussed by Bruck [2, 3], Bruck and Bose [4, 5]. However, the results here could be used to extend some of the covering properties of existing spreads to cases with larger N, m, t.

Acknowledgment

This paper has been extracted from the author's Ph.D. dissertation, (University of North Carolina, 1967), the research for which was directed by R. C. Bose.

References

1. R. C. Bose and G. G. Koch, The Design of Combinatorial Information Retrieval Systems for Files with Multiple-Valued Attributes, University of North Carolina, Institute of Statistics Mimeo Series No. 552, October 1967, to appear in SIAM.
2. R. H. Bruck, Recent Advances in the Foundations of Euclidean Plane Geometry, *Amer. Math. Monthly* **62**, No. 7, II, Slaught Memorial Paper No. 4 (1955).
3. R. H. Bruck, Existence Problems for Classes of Finite Planes, Mimeographed notes of lectures delivered to the Seminar of the Candian Mathematical Congress, meeting in Saskatoon, Sask., August 1963.

4. R. H. BRUCK AND R. C. BOSE, The Construction of Translation Planes from Projective Spaces, *J. Algebra* **1** (April 1964), No. 1.
5. R. H. BRUCK AND R. C. BOSE, Linear Representations of Projective Planes in Projective Spaces, *J. Algebra* **4** (July 1966), No. 1.
6. D. K. RAY-CHAUDHURI, Combinatorial Information Retrieval Systems for Files, *IBM Res. Rept. RC*-1554, IBM Watson Research Center, Yorktown Heights, N.Y., February 1966, to appear in SIAM.